

Electronics and
Computer Science



PSIGate

JISC

eBank UK:

Dissemination of research data using EPrints

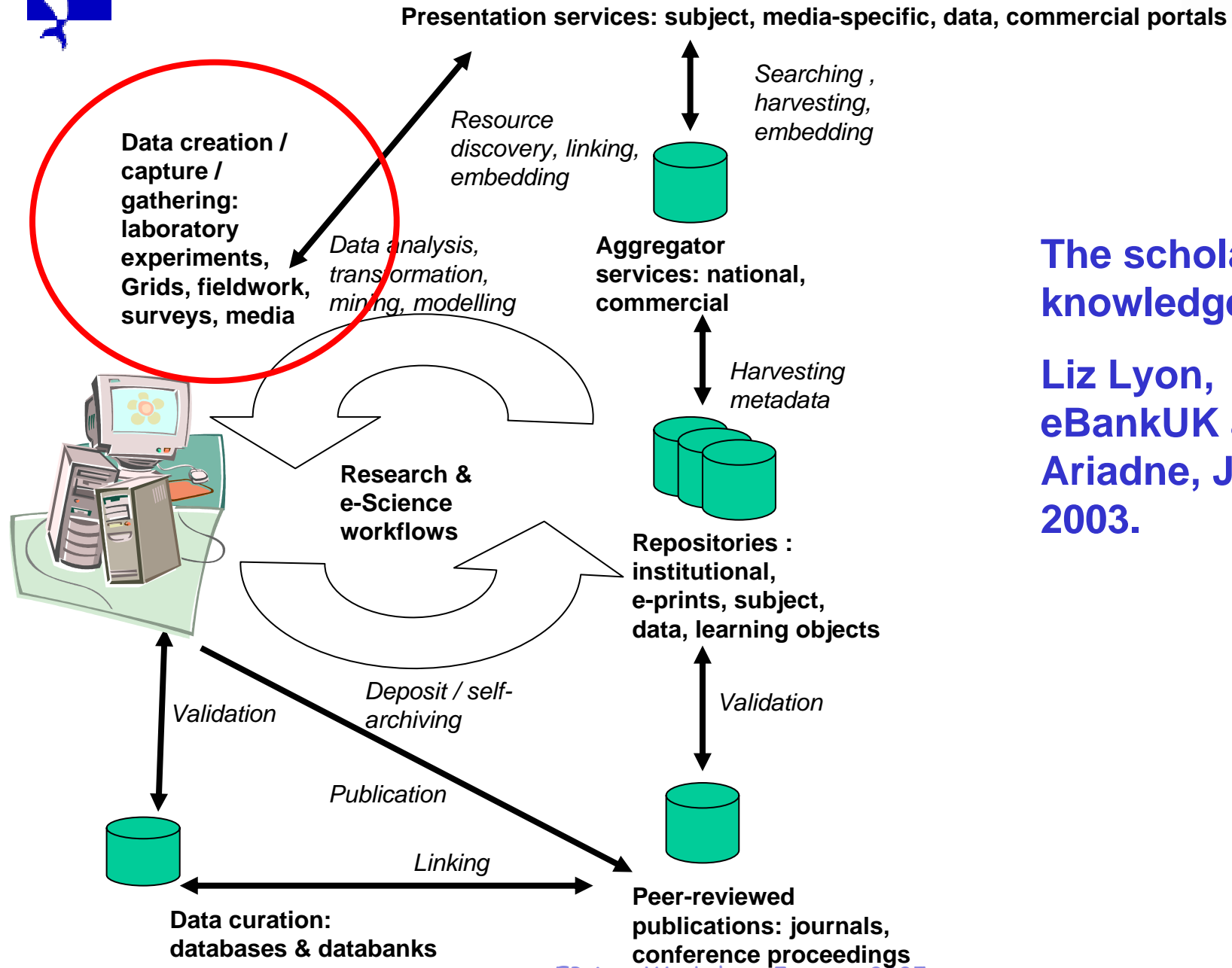
Simon Coles, School of Chemistry, University of Southampton



Overview

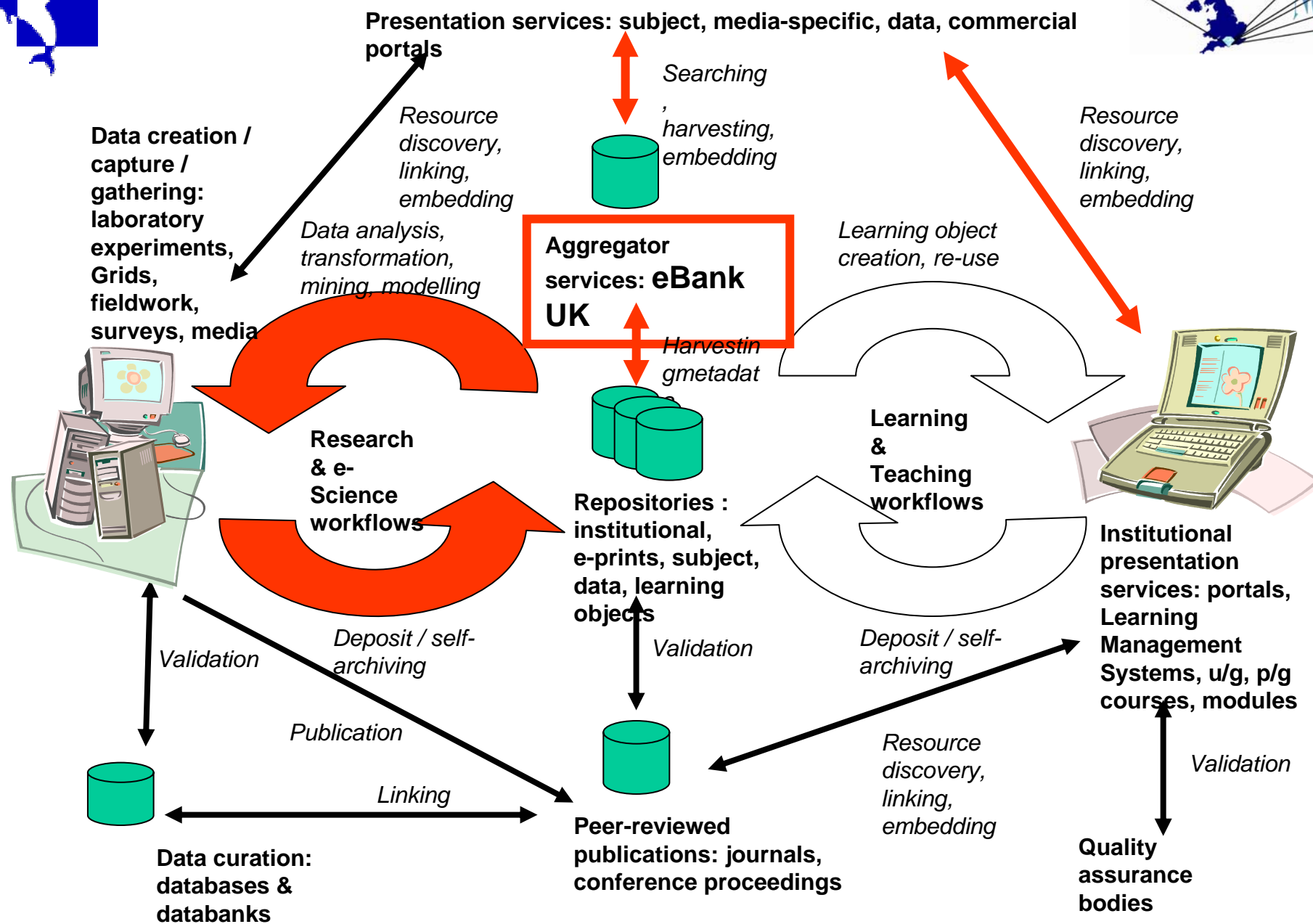


- Scholarly communications in Chemistry
 - Data, information, workflows and provenance
- The data publication bottleneck
 - e-Science and chemistry
- eBank UK
 - Information architecture, data flow and interoperability
- Challenges for the future
 - Expansion into other disciplines and data formats



The scholarly knowledge cycle.

**Liz Lyon,
eBankUK article.
Ariadne, July
2003.**



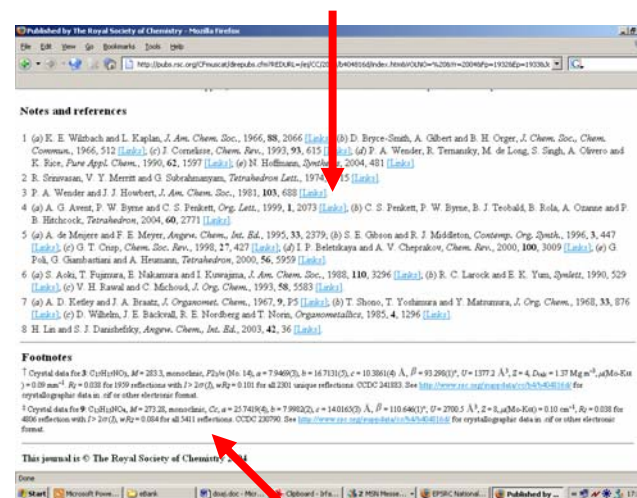
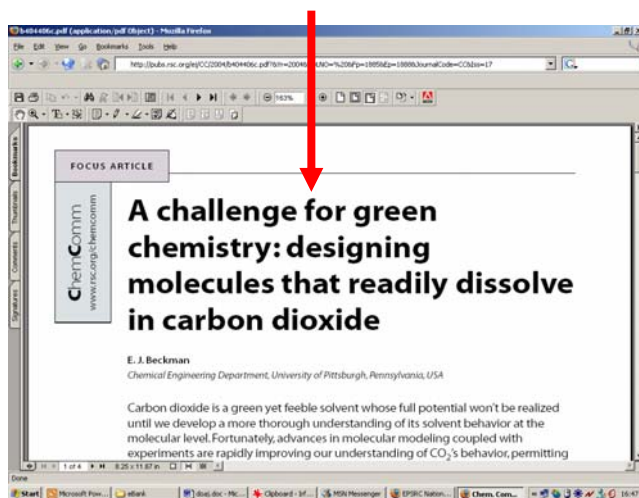


Current chemistry publishing protocols

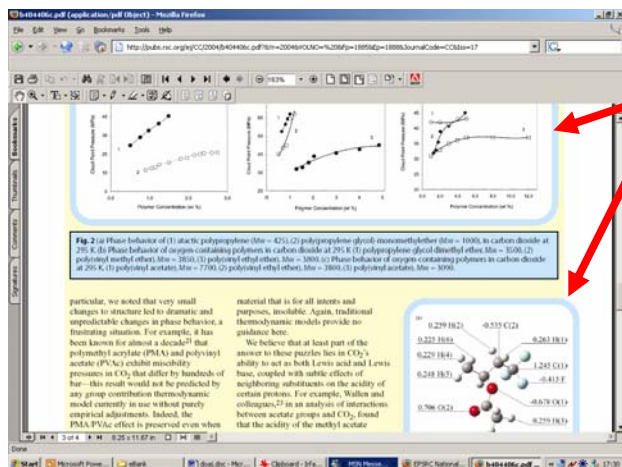


Ideas and interpretations

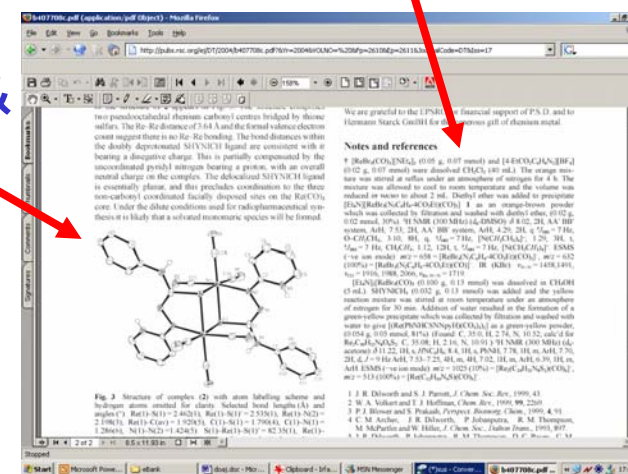
Hooks into the literature



Raw data!



Results & derived data





Directory of open access journals - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://www.doaj.org/ljbs?cpid=61

[Chinese Journal of Chemistry](#)
ISSN: 1001604x
Subject: [Chemistry \(General\)](#) --- [Analytical Chemistry](#) --- [Inorganic Chemistry](#) --- [Organic Chemistry](#)
Publisher: The Chinese Chemical Society and Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences
Language: English, Chinese
Keywords: chemistry
Start Year: 2000

[Croatia Chemica Acta](#)
ISSN: 1334417x
Subject: [Chemistry \(General\)](#)
Publisher: Croatian Chemical Society
Language: English
Keywords: chemistry
Start Year: 1998

[Current Science](#)
ISSN: 00113891
Subject: [Chemistry \(General\)](#) --- [Biology](#) --- [Environmental Sciences](#)
Publisher: Current Science Association ; Indian Academy of Sciences
Language: English
Keywords: chemistry, science, environmental science, biology
Start Year: 1998

[Eclética Química](#) **DOAJ Content**
ISSN: 01004670
EISSN: 16784618
Subject: [Chemistry \(General\)](#)
Publisher: Fundação Editora da Universidade Estadual Paulista Júlio de Mesquita Filho - UNESP
Language: English, Portuguese
Keywords: Exact and Earth Sciences

Done

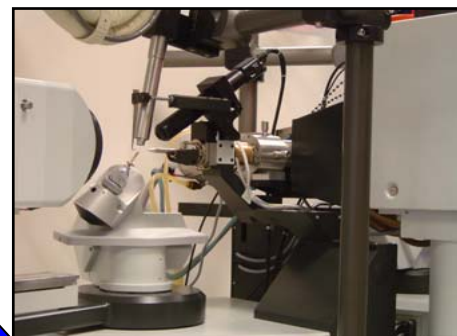
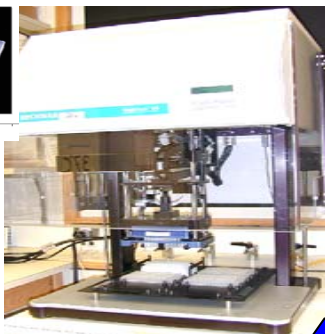
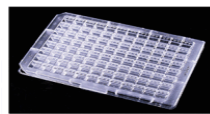
Start Directory of open acc... MSN Messenger Document1 - Microsoft ... 12:09

Done

Start Directory of open acc... MSN Messenger Document1 - Microsoft ... 12:07



The data deluge



**Data
Overload!**



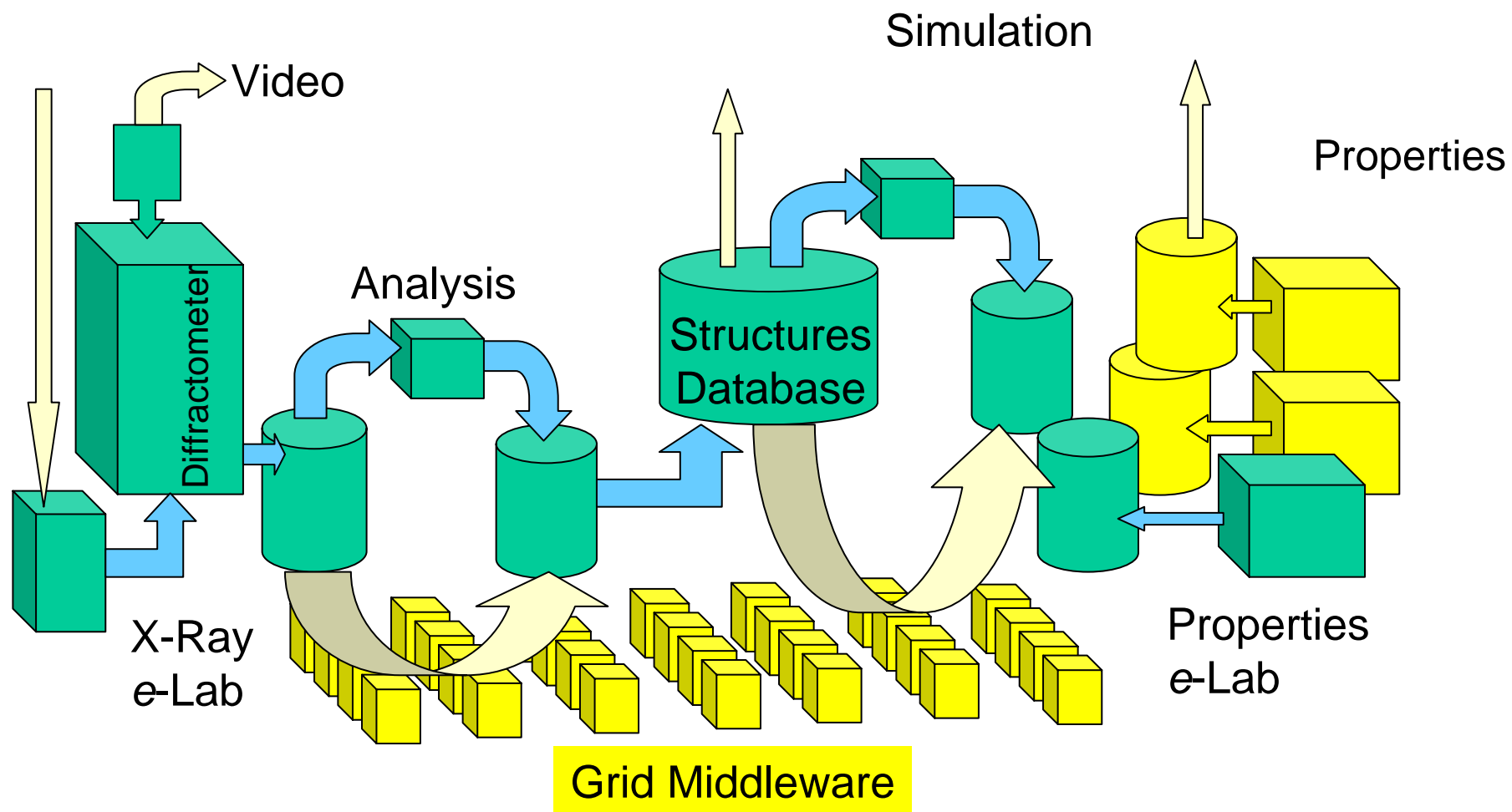
EPSRC National
Crystallography
Service



**How do we
disseminate?**



CombeChem: eScience testbed





Establishing common ground...



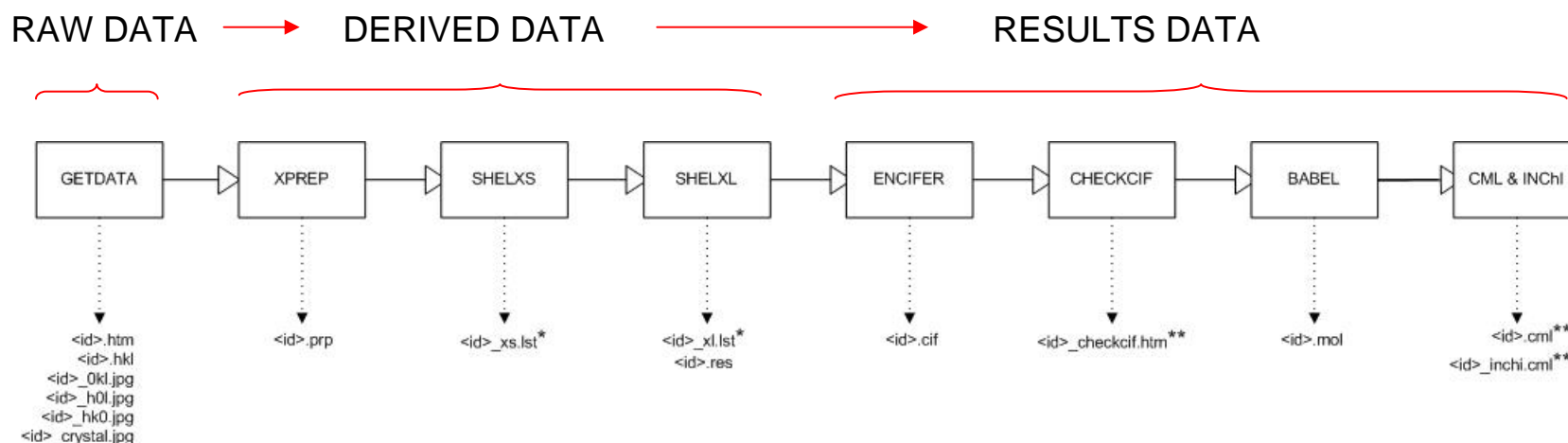
- Understand the data creation process
- Terminology and definitions
 - Data
 - Metadata
 - Datafile
 - Dataset
 - Data holding
- Different views
 - Digital library researchers, computer scientists, chemists
 - Generic vs specific
 - Modeller vs practitioner
- Aim for a common ontology
- Modelling the domain
- Creating a metadata schema



Crystallography workflow



- **Initialisation:** mount new sample on diffractometer & set up data collection
- **Collection:** collect data
- **Processing:** process and correct images
- **Solution:** solve structures
- **Refinement:** refine structure
- **CIF:** produce CIF (Crystallographic Information File format)
- **Report:** generate Crystal Structure Report





Deposition into the archive



Southampton Crystal Reports - Core Bibliographic Information - Mozilla Firefox

http://ncrystals.chem.soton.ac.uk/per/users/submit?_action_new=1

University of Southampton Crystal Structure Report Archive

Home - About - Browse - Search - Register - User Area - Help

Core Bibliographic Information

Please enter the bibliographic data about your deposit. Fields marked with a * are fields that must be filled out before your deposit will be accepted.

< Previous Save for Later Next >

Title of Compound *
The title of the Compound
Example: N-5-Chloro-pyridin-2-yl-4-fluoro-benzamide

Authors/Creators *
Enter in the format given on the first page or title page. If there are more than four authors, click on the [More Spaces] button.
Example: [Smith] [J.P.] [j.smith@soton.ac.uk]
Example: [Chance-Collins] [Chris] [chris@bar.com]
Example: [Fu] [Yan-Li] []
Example: [von Hopffner] [Ludwig G.] []

Family Name	Given Name / Initials	Creators email (if known)
1 Coles	Simon J	s.i.coles@soton.ac.uk

Southampton Crystal Reports - Core Bibliographic Information - Mozilla Firefox

http://ncrystals.chem.soton.ac.uk/per/users/submit?_action_new=1

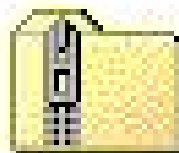
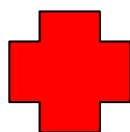
Institution *
Please enter the institution from which the item originated.
Example: University of Whitezombia
University of Southampton

Compound Class *
Organic
Bio-Organic
Organometallic
Inorganic

Empirical Formula *
The formula of the compound. Numbers will be displayed as subscript, but not while you are inputting them.

Uncontrolled Keywords
Natural language terms to describe the content of the item

ICHI Code *
This code can be obtained from the Conversion Server at Cambridge.



id.zip



An Archive entry



EBank Southampton - 2-(N-Ferrocenylmethylcarbamoyl)-5-(methoxycarbonyl)-3,4-diphenylpyrrole - Mozilla Firefox

File Edit View Go Bookmarks Tools Help

http://eprints.ebank.ecs.soton.ac.uk/archive/00000025/

University of Southampton Crystal Structure Report Archive

Home - About - Browse - Search - Register - User Area - Help

2-(N-Ferrocenylmethylcarbamoyl)-5-(methoxycarbonyl)-3,4-diphenylpyrrole

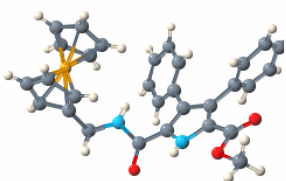
Susanne L. Huth, Michael B. Hursthouse, Simon J. Coles, Mark E. Light, Peter N. Horton, Phil A. Gale, G. Denuault and C. N. Warriner.

University of Southampton

C30H26FeN2O3

CCDC Code: XUZSUG
ICHI Code: C30H26FeN2O3, 1H3-36-20(35)18-16(14-10H-6H-3H-7H-11H-14)15(13-8H-4H-2H-5H-9H-13)17(33H-18)19(34)32H-12H2-30-28H-26H-27H-29H(30)31(26,27,28,30)21H-22H(31)24H(31)25H(31)23H(21)31
Compound Class: Organometallic
Keywords: Supramolecular Chemistry
Creation Date: 31 May 2002
Deposited By: Susanne L. Huth
Deposited On: 10 May 2004

Data collection parameters



Applet jmol started

Start Pegasus Mail WinGX v1.6... ISSC-XIII - M... EBank Sout... Windows Mes... eBank ISSC a... ECM22EPrint..

ecrystals.chem.soton.ac.uk

Data collection parameters

Chemical formula	C30 H26 Fe N2 O3
Crystallisation Solvent	
Crystal morphology	
Crystal system	Orthorhombic
Space group symbol	Pbca
Cell length a	6.0816(4)
Cell length b	24.8503(16)
Cell length c	31.120(3)
Cell angle alpha	90.00
Cell angle beta	90.00
Cell angle gamma	90.00
Data collection temperature	120(2)

Refinement results

Solution figure of merit	
R Factor (Obs)	0.0573
R Factor (All)	0.1185
Weighted R Factor (Obs)	0.1046
Weighted R Factor (All)	0.1243

Available Files

Final Result

02sot064.CIF	19k
02sot064.cml	8k
02sot064_checkcif.html	14k

Refinement

02sot064.RES	9k
--------------	----

Solution

02sot064.PRZ	5k
--------------	----

Processing

02SOT064.HTM	6k
02sot064.HKL	338k

Other Files

02sot064.DOC	113k
02sot064.LST	49k



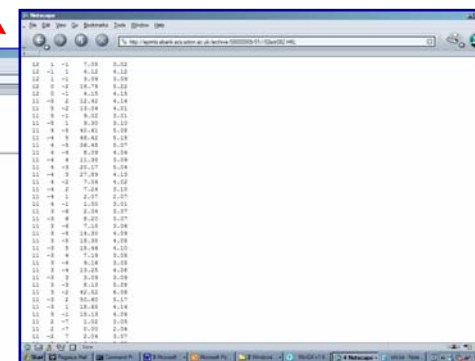
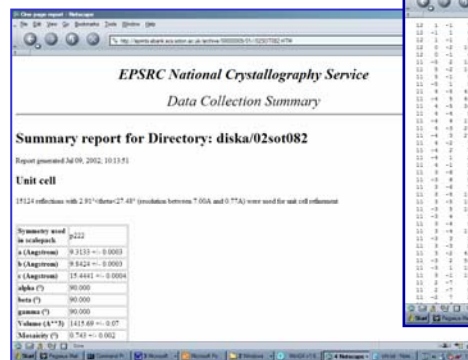
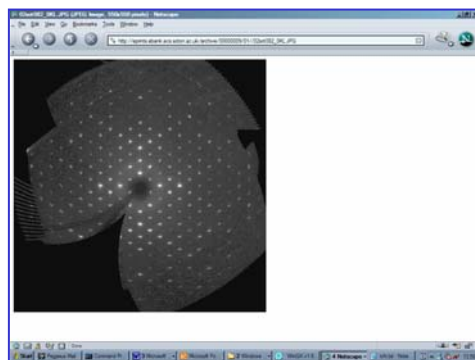
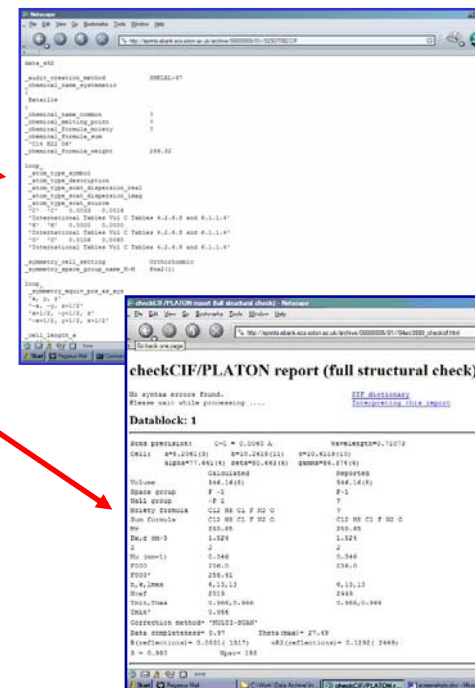
Access to the underlying data



Data collection parameters	
Chemical formula	C30 H26 Fe N2 O3
Crystallisation Solvent	
Crystal morphology	
Crystal system	Orthorhombic
Space group symbol	Pbca
Cell length a	6.0816(4)
Cell length b	24.8503(16)
Cell length c	31.120(3)
Cell angle alpha	90.00
Cell angle beta	90.00
Cell angle gamma	90.00
Data collection temperature	120(2)

Refinement results	
Solution figure of merit	
R Factor (Obs)	0.0573
R Factor (All)	0.1185
Weighted R Factor (Obs)	0.1046
Weighted R Factor (All)	0.1243

Available Files	
Final Result	
02sot064.CIF	19k
02sot064.cml	8k
02sot064_checkcif.html	14k
Refinement	
02sot064.RES	9k
Solution	
02sot064.PRP	5k
Processing	
02SOT064.HTM	6k
02sot064.HKL	338k
Other Files	
02sot064.DOC	113k
02sot064.LST	49k

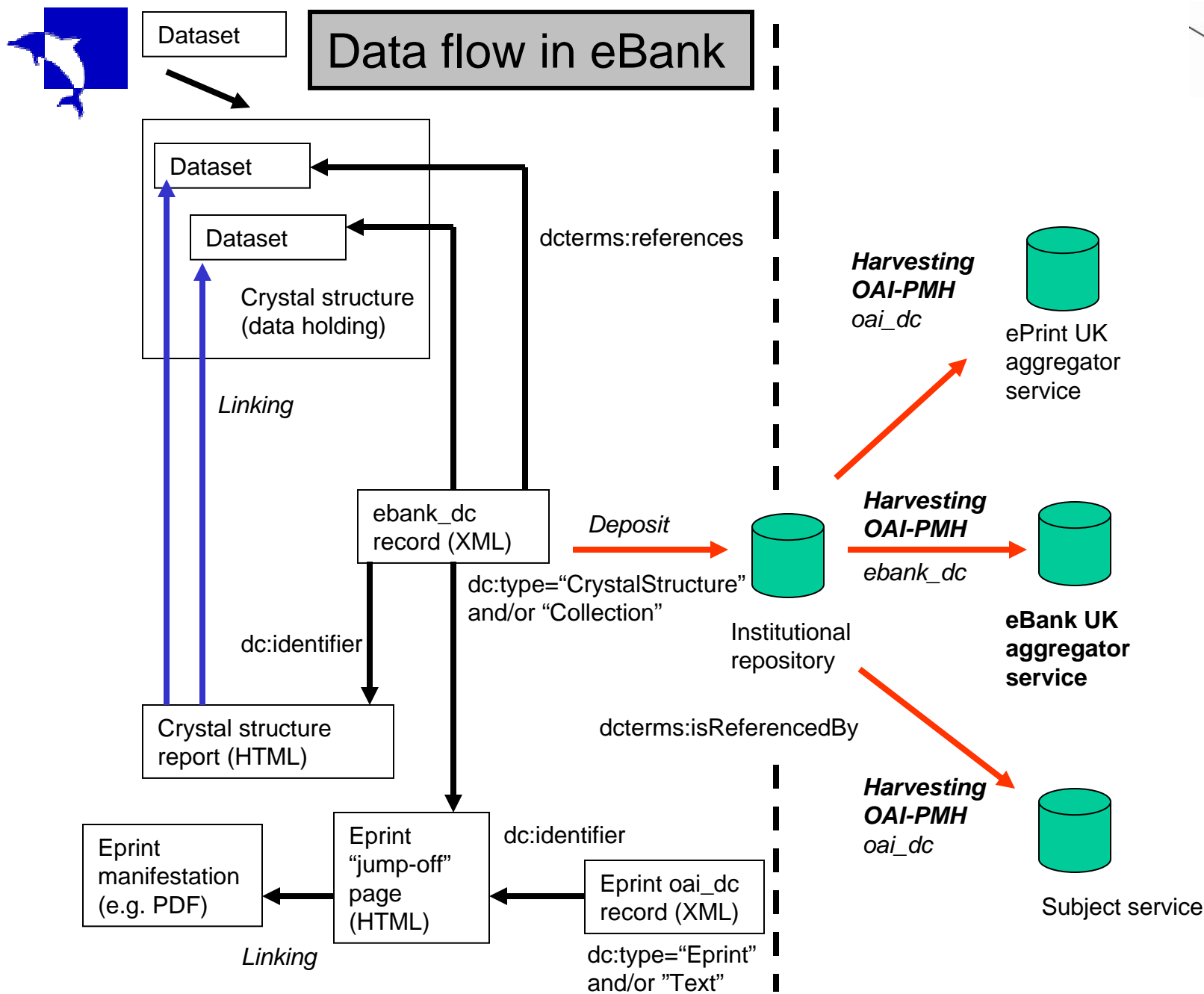




Some metadata issues



- Using simple and qualified Dublin Core
- Additional chemical information in schema for harvesting e.g. empirical formula
- Schema contains International Chemical Identifier (InChI)
- Links to all datasets associated with an experiment
- Links to individual datasets *within* an experiment
- Links to EPrints (and other published literature) derived from the data
- Using vocabularies specific to crystallography
- Engaging the broader scientific community to ensure different schemas are compliant and standards can emerge





Harvesting: OAIster



OAIster Search - Mozilla Firefox

http://oaiSTER.unid.umich.edu/cgi/bib/bib-idx?c=oaiSTER.page=simple

Home Search View Institutions Help

Use this form to find digital resources (what is a digital resource?) from many institutions (which institutions are available?).

Why do I sometimes get duplicate records? Why do I sometimes have trouble accessing digital objects? Check out this explanation.

See help for search tips:

Search all fields (help)
Use a word or phrase, e.g., diploma*, fancy dress

and P or C

Search within particular fields (help)
Use a word or phrase, e.g., diploma*, fancy dress

Title: and
Author/Creator: coles and
Subject: crystallography and
Resource Type: all types

Choose sorting of results (help)
e.g., results in date descending order, from 2002 to 1999

☒ title ☐ author/creator
☐ date descending ☐ date ascending

Done

OAIster Search Results - Mozilla Firefox

http://oaiSTER.unid.umich.edu/cgi/bib/bib-idx?type=boolean&size=10&c=oaiSTER&sid=f5901b99a0e5976b4012823cf2c36c28q1=8rgn1

Jump to Records: 1 | 11 | 21 | 31 | 41 | 51 | 61 | 71 | 81 | 91 | 101 | 111 | 121 | 131 | 141 | 151 | 161 | 171 | 181 | 191 | 201 | 211 | 221 | 231 | 241 | 251 | 261 | 271 | 281 | 291 | 301 | 311 | 321 | 331 | 341 | 351 | 361

Record 1 of 365

Title	Author/Creator
1,1'-bis(((6-Propionylaminopyrid-2-yl)amino)carbonyl)ferrocene	Westwood, J.
	Coles, S. J.
	Collinson, S. R.
	Gasser, G.
	Green, S. J.
	Hursthouse, M. B.
	Light, M. E.
	Tucker, J. H. R.
Year	2000-10-26
Resource Type	Crystal Structure
Resource Type	Completed Analysis

Note: A series of ferrocene-containing amidopyridyl receptors, in two regioisomeric forms (1,1' and 1,3), have been shown to bind cyclic organic molecules in chloroform through complementary hydrogen-bonding interactions. Complexation was monitored by NMR spectroscopy and by cyclic voltammetry. The magnitude of the host-guest binding strength and the redox response to complexation depend on both the type and the relative position of the amidopyridyl groups on the cyclopentadienyl rings of the ferrocene.

Note: http://ecrystals.chem.soton.ac.uk/archive/00000083/01/00sro475_data/00sro475_checkcif.html

Note: http://ecrystals.chem.soton.ac.uk/archive/00000083/01/00sro475_data/00sro475.cml

Note: http://ecrystals.chem.soton.ac.uk/archive/00000083/01/00sro475_data/00sro475.mol

Note: http://ecrystals.chem.soton.ac.uk/archive/00000083/01/00sro475_data/00sro475.res

Note: http://ecrystals.chem.soton.ac.uk/archive/00000083/01/00sro475_data/extracted.cml

Note: http://ecrystals.chem.soton.ac.uk/archive/00000083/01/00sro475_data/00sro475_xl.lst

Note: http://ecrystals.chem.soton.ac.uk/archive/00000083/01/00sro475_data/00sro475.HKL

Note: http://ecrystals.chem.soton.ac.uk/archive/00000083/01/00sro475_data/00sro475_xs.lst

Done

Start **OAIster Search Resu...** **MSN Messenger** **Document1 - Microsoft ...**



Linking and aggregating



eBank UK Demo

This is a prototype interface for the [eBank UK](#) JISC-funded project. It demonstrates an OAI-PMH aggregator service which cross searches a small sample of metadata records describing crystallography experiments (provided by the National Crystallography Service at the University of Southampton), and a small number of metadata records describing articles from the Crystallography literature (made available for use in this demo only by IUCr). Links to the crystallography data sets and to the articles on line at the IUCr website are available in the search results.

Search for entries matching all the following:

Author =
CCDC Code =
IUPAC name =
Empirical Formula =
Compound Class =
General keywords =
Date released =
OR published in the last

Search within: ☒ Data Reports ☒ Publications e.g. journal articles

eBank UK Demo

Crystal Structure Data Reports

[Crystal Structure Report of 2-\(N-Ferrocenylmethylcarbamoyl\)-5-\(N-phenylcarbamoyl\)-3,4-diphenylpyrrole](#)
Creator(s): Hursthouse, Michael B., Light, Mark E., Coles, Simon J., Horton, Peter N., Gale, Phil A., Denuault, G., Warriner, C. N.
Date released: 23/05/2004
Empirical Formula: C₃₅H₂₉FeN₃O₂
IUPAC name: 2-(N-Ferrocenylmethylcarbamoyl)-5-(N-phenylcarbamoyl)-3,4-diphenyl pyrrole
CCDC code: XU2SIU
Compound Class: Organic
General keywords: Supramolecular Chemistry
Related article: [2A URI citation?](#)

Available Datafiles

CIF file
processing Dataset
refinement Dataset
solution Dataset

[Crystal Structure Report of 2-\(N-Ferrocenylcarbamoyl\)-5-\(methoxycarbonyl\)-3,4-diphenylpyrrole](#)
Creator(s): Hursthouse, Michael B., Coles, Simon J., Light, Mark E., Horton, Peter N., Gale, Phil A., Denuault, G., Warriner, C. N.
Date released: 23/05/2004
Empirical Formula: C₂₉H₂₄FeN₂O₃
IUPAC name: 2-(N-Ferrocenylcarbamoyl)-5-(methoxycarbonyl)-3,4-diphenylpyrrole
CCDC code: XU2SOA
Compound Class: Organometallic
General keywords: Supramolecular Chemistry
Related article: [2A URI citation?](#)

Available Datafiles

CIF file
processing Dataset
refinement Dataset
solution Dataset

Publications

A supramolecular assembly: aquatris(pentafluorophenyl)borane as its mixed dimethyl sulfone and water solvate, (H₂O)₃(C₆F₅)₃Me₂SO₂H₂O

The title compound, C₁₈H₂BF₆O₄S₂, obtained by crystallization of a product formed from a reaction mixture containing B(C₆F₅)₃ and Me₂SO₂ (and H₂O) in hexane, was characterized in the solid state as a supramolecular assembly containing water adducts of tris(pentafluorophenyl)borane, (H₂O)₃(C₆F₅)₃, linked together by a network of hydrogen bonds involving one additional H₂O and one additional Me₂SO₂ molecule per adduct molecule.

Creator(s): Coles, Simon J., Hursthouse, Michael B., Beckett, Michael A., Dutton, Michael
Acta Crystallogr E Struct Rep Online Vol 59 Issue Pt 9 pp. o1354 - o1356
DOI: [10.1107/S1600536802004786](#)
Download from: [http://scripts.iucr.org/cgi-bin/getarticleid?asin=1600-5368&volume=58&page=o1354&detail=yes](#)

Structural investigations of phosphorus-nitrogen compounds. 5. Relationships between molecular parameters of 2,2-diphenyl-4,6-cis-oxyltetra(ethyleneoxy)-4,6-R₂-cyclotriphosphazatrienes (R = Cl, OCH₂CF₃, OPh, OMe, NHPH, NHBut) and substituent basicity constants

The syntheses and crystal structures of six new cis-ansa derivatives N3P3Ph2O(CH₂CH₂O)₄ R₂ (R = Cl, OCH₂CF₃, OPh, OMe, NHPH, NHBut) are reported and the observed relationship between molecular parameters of the N3P3 ring and substituent basicity constants is discussed.

Creator(s): Besli, S., Coles, S. J., Hursthouse, M. B., Kilic, A., Mayer, T. A., Shaw, R. A.
Acta Crystallogr B Vol 58 Pt 6 pp. 1067 - 1073
DOI: [10.1107/S0108768102018608](#)
Download from: [http://scripts.iucr.org/cgi-bin/getarticleid?asin=0108-7681&volume=58&page=1067&detail=yes](#)

Related dataset: [http://ccrystals.chem.soton.ac.uk/archives/00000062/](#)

5alpha-Cholestane

The title compound, C₂₇H₄₈, is a steroid derivative composed of a saturated-carbon fused-ring framework with two methyl substituents and an alkyl side chain.

Creator(s): Coles, S. J., Hursthouse, M. B., Frampton, C. S.
Acta Crystallogr E Struct Rep Online Vol 58 Issue Pt 4 pp. o445 - o446
DOI: [10.1107/S1600536802004786](#)
Download from: [http://scripts.iucr.org/cgi-bin/getarticleid?asin=1600-5368&volume=58&page=o445&detail=yes](#)

Related dataset: [http://ccrystals.chem.soton.ac.uk/archives/00000051/](#)

Ethyl (2S)-2-[(2R*,2R*)-2'-[5-(dimethyl-5'-oxopenthydryl-2,2'-bifuranyl-5-yl)-2'-hydroxyethanoate

The framework of K₂Zn(H₂P₂O₇)₂·2H₂O contains acid diphosphate-metalate layers linked by KO interactions and weak hydrogen bonds. Zn²⁺ cations are coordinated octahedrally by O atoms from two bidentate [H₂P₂O₇]²⁻ anions and two water molecules.



Embedded in a science portal





PSigate
Physical Sciences Information Gateway

SEARCH SUBJECTS

- Astronomy
- Chemistry
- Earth sciences
- Materials science
- Physics
- History & policy

FEATURES

- PSigate home
- Search PSigate
- Subject Headings A-Z
- About PSigate
- Site map
- New additions
- Feedback form
- Suggest site
- News services
- Reference
- Spotlight
- Science Timelines
- Hot Topics
- Science Data
- Courses
- Jobs
- Learning & Teaching

PSigate Home > eBank

This is a prototype test interface to the eBank UK service providing access to data in the University of Southampton eCrystal data repository and elsewhere. eBank UK is a JISC-funded project which is a part of the Semantic Grid Programme. The project is being led by UKOLN in partnership with the Combechem project at the University of Southampton and PSigate.

Enter your search term(s)

Author

CCDC Code

IUPAC name

Empirical Formula

Compound Class

General keywords

Date released

OR published in the last

Search within ☒ Data Reports ☒ Publications e.g. journal articles

Copyright © PSigate. All rights reserved. PSigate is a service of the Resource Discovery Network (RDN). feedback@psigate.ac.uk



PSigate
Physical Sciences Information Gateway

About Us | Contacts | Site Map | Help

SEARCH SUBJECTS

- Astronomy
- Chemistry
- Earth sciences
- Materials science
- Physics
- History & policy

FEATURES

- PSigate home
- Search PSigate
- Subject Headings A-Z
- About PSigate
- Site map
- New additions
- Feedback form
- Suggest site
- News services
- Reference
- Spotlight
- Science Timelines
- Hot Topics
- Science Data
- Courses
- Jobs
- Learning & Teaching

PSigate Home > eBank > Search Results

Your search returned 28 data reports and 4 publications. Viewing 1 to 10

[Next](#)
[New search](#)

Crystal Structure Data Reports

Crystal Structure Report of 2-(N-Ferrocenylmethylcarbamoyl)-5-(N-phenylcarbamoyl)-3,4-diphenyl pyrrole

Creator(s): Hursthouse, Michael B., Light, Mark E., Coles, Simon J., Horton, Peter N., Gale, Phil A., Denuault, G., Warriner, C. N.

Date released: 23/05/2004

Empirical Formula: C35H29FeN3O2

IUPAC name: 2-(N-Ferrocenylmethylcarbamoyl)-5-(N-phenylcarbamoyl)-3,4-diphenyl pyrrole

Compound Class: Organic

General keywords: Supramolecular Chemistry

Related article: [2A URI citation?](#)



Current situation



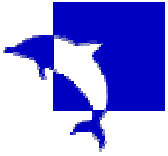
- Version 2.0 eBank metadata schema
- Pilot institutional e-data repository for harvesting (raw, derived, results data) using EPrints.org software
- Exports records as ebank_dc and oai_dc
- Validation of schema & discussion with International Union of Crystallography for final developments and wider deployment
- Pilot eBank UK aggregator service
- Developing search interface Version 1.0
- Testing with PSIgate physical sciences portal – embedding eBank UK



What's next?



- Progress towards generic metadata schemas
- Validation against other schema (CCLRC Model)
- Eprints.org software: allow for more generic scientific data and schemas?
- Metadata enhancement: keywords based on knowledge of keywords in related publications?
- Investigate identifiers: International Chemical Identifier
- Explore context sensitive linking
- Full embedding into chemical and crystallographic research and publishing
- e-Learning embedding and pedagogic evaluation
- Feasibility study in related domains



Breakout Session?



- Describing non 'Dublin Core' terms
 - Qualified Dublin Core
 - Complex object formats: METS vs MPEG-21 DIDL
 - Set & Friends containers
- Compliance between schemas
 - One generic schema
 - Develop multiple schemas
- Rights
 - Use / reuse
 - Publisher
- Linking & aggregating
 - DOI
 - Keyword ontologies
 - Identifiers
 - Context sensitive linking